

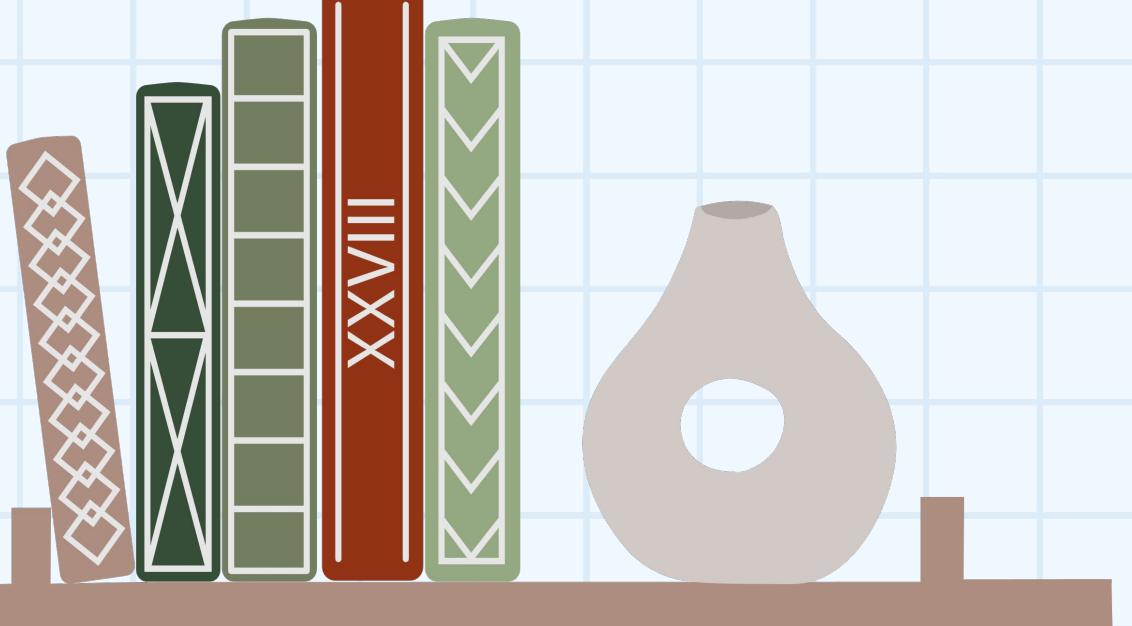
Powered by
Futurense



BS./BSC.IN

Applied ai and Data Science

Basics of Data Analytics





Powered by

Futurense

Let's dive into and learn:



1

Types of Data Visualizations

Data Visualization



Powered by

Futurense

- Data Visualization is the science and art of communicating information through the use of visual elements such as graphs, charts, and maps.
- Its goal is to make information easy to comprehend, interpret, and retain.
- Using data visualization, you can provide a summary of the relationships, trends, and patterns hidden in the data

Data Visualization



Powered by

Futurense

- Need to choose the visualization that most effectively delivers your findings to your audience.
- Ask yourself the following questions
 - What am I trying to show?
 - What do I want to compare?
 - What relationships do I want to show?
 - Do I want my audience to interact with the data?

Data Visualization



Powered by

Futurense

- Are you trying to show the different values of a particular feature?
 - E.g. Distribution of total values of transaction?
- Are you trying to show sub-parts or components of a whole?
 - E.g. Contribution of different product lines in the total revenue of the company
- Are you trying to show the evolution over time?
 - E.g. Revenue of a company over the past five years?

Data Visualization



Powered by

Futurense

- Univariate Visualization
 - One variable
 - Explore the statistical distribution of each variable individually
 - Distribution – how the data is spread across different values
- Bivariate visualization
 - Two variables
 - Explore the relationship between two variables
- Can be expanded to more variables using other aspects of

Univariate Visualization



Powered by

Futurense

- The first step in any comprehensive data analysis is to explore each variable
 - Can be continuous (e.g. age, weight, income)
 - Can be categorical (e.g. gender, state of residence)

Histogram



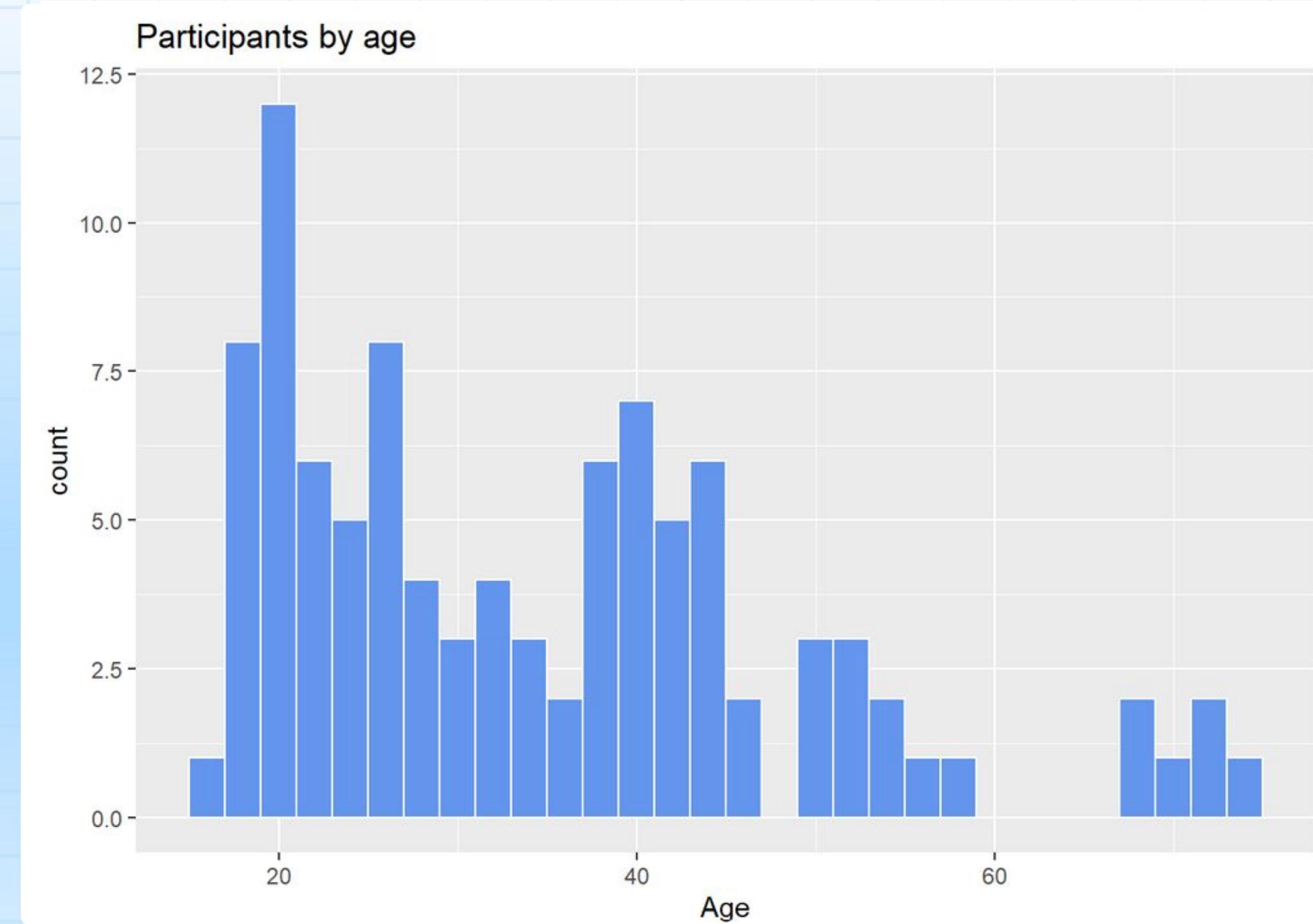
- Histograms are the most common approach to visualizing a quantitative variable.
- In a histogram, the values of a variable are typically divided up into adjacent, equal width ranges called bins
- The number of observations in each bin is plotted with a vertical bar.

Histogram



Powered by

Futurense



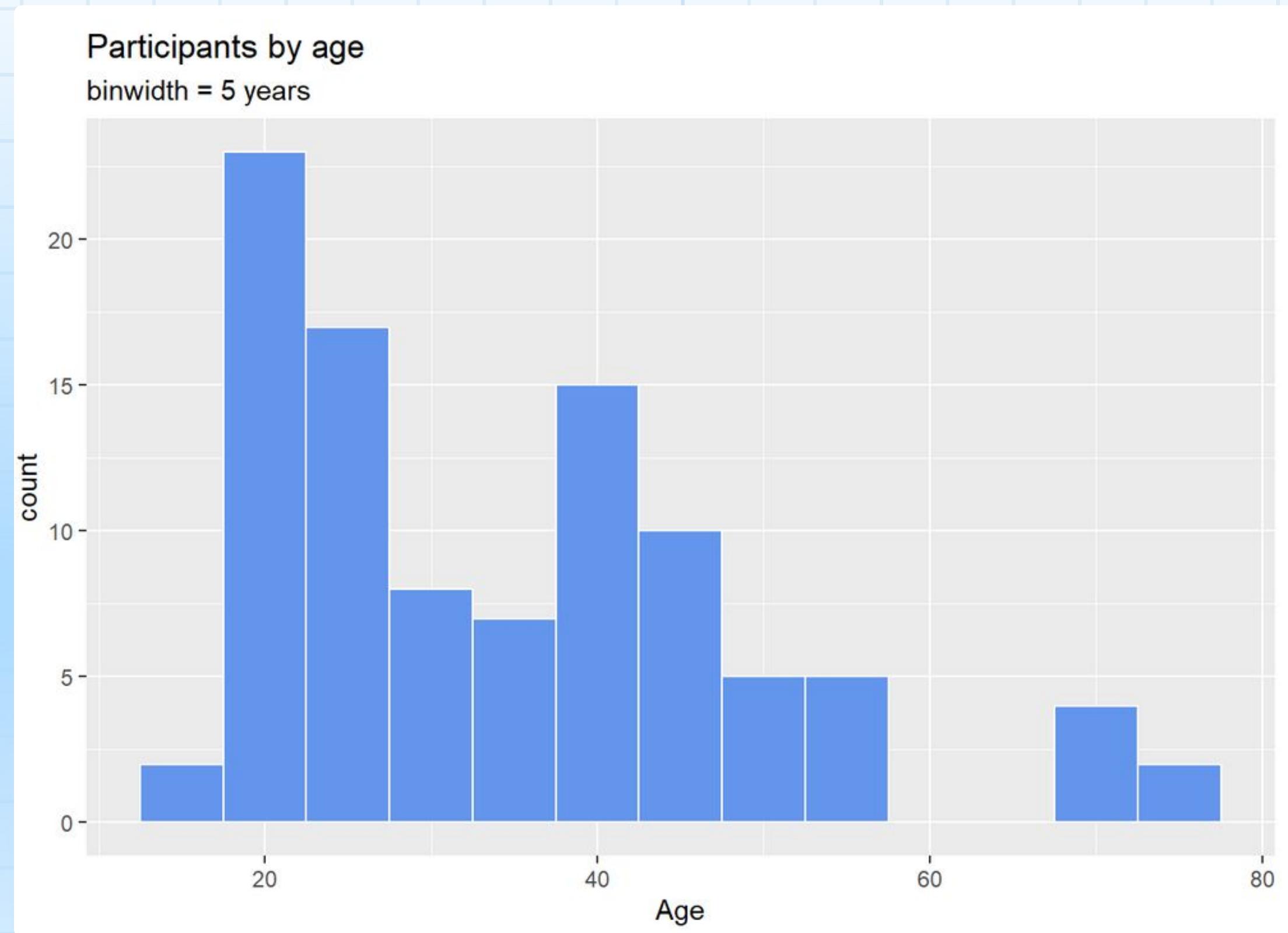
Shows the distribution
of a particular variable

Histogram



Powered by

Futurense



BINWIDTH

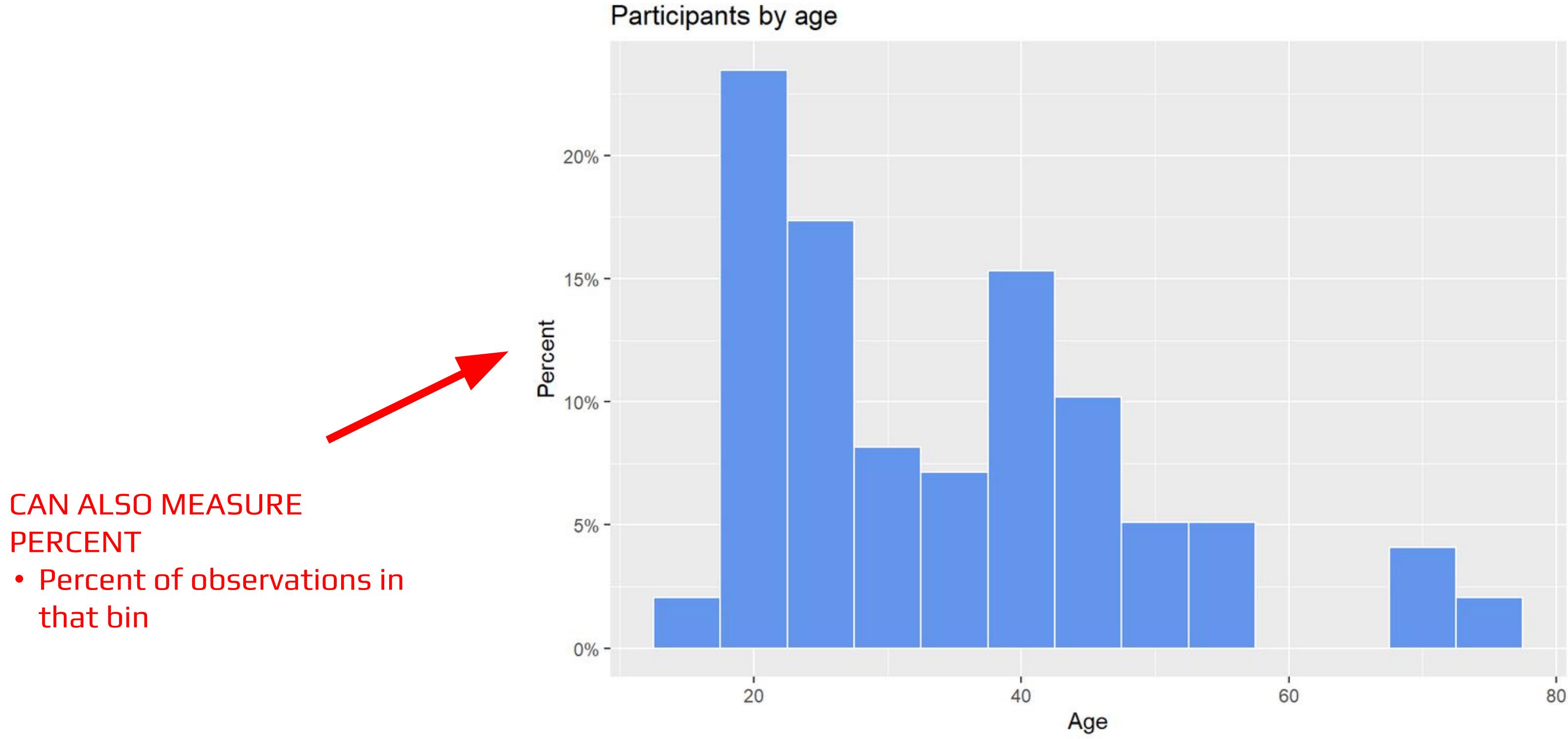
- Most important choice for histogram
- Bin size can reveal or hide patterns in data
- Large bins will include many points in each bin. Can hide patterns
- Small bins will include few points in each bin. Can reveal patterns

Histogram



Powered by

Futurense





Categorical Variables

- The distribution of a categorical variable can be illustrated by showing the count or frequency or percentage of records in each category.

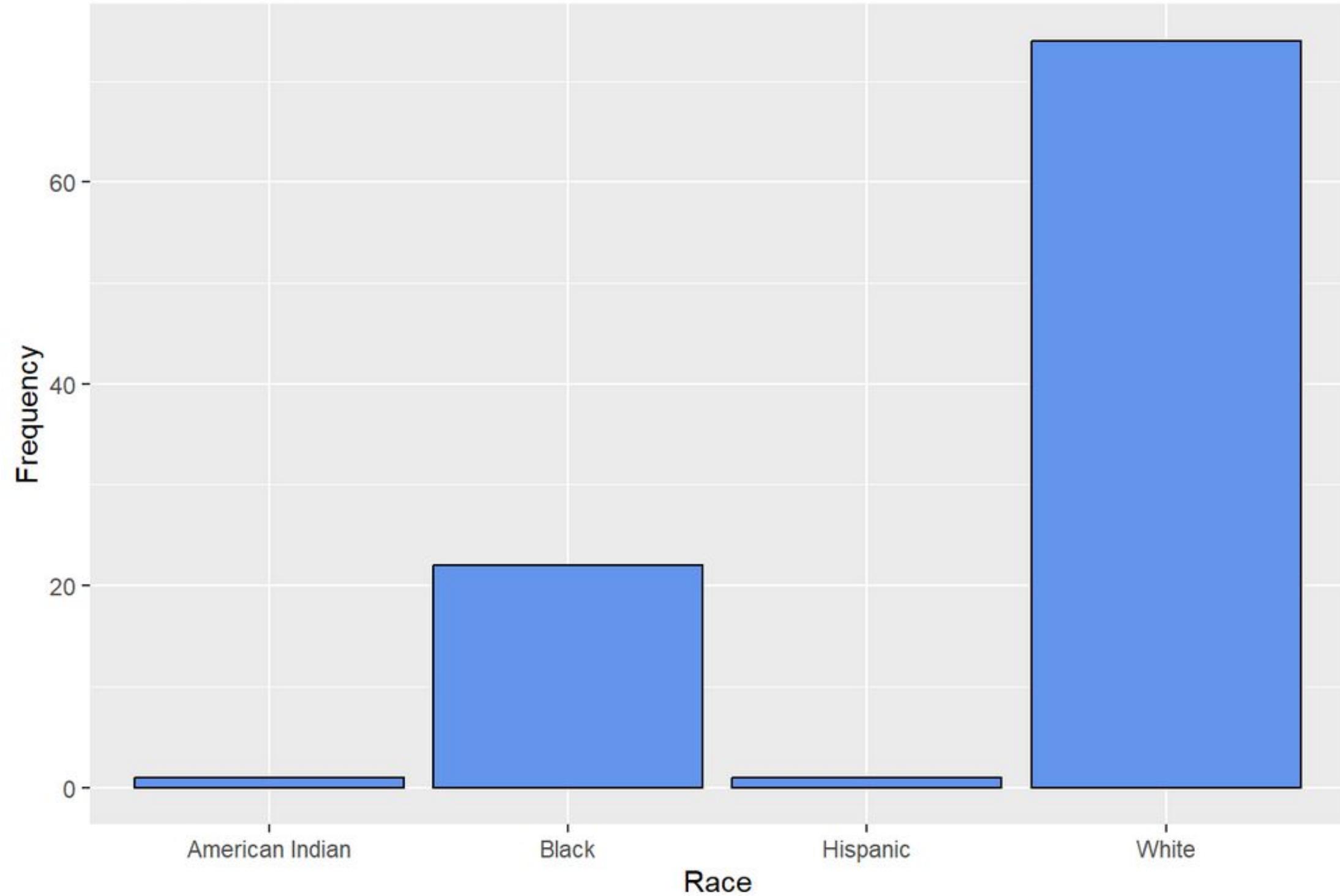
Bar Graph



Powered by

Futurense

Participants by race



The height of the bar shows the frequency of each category.

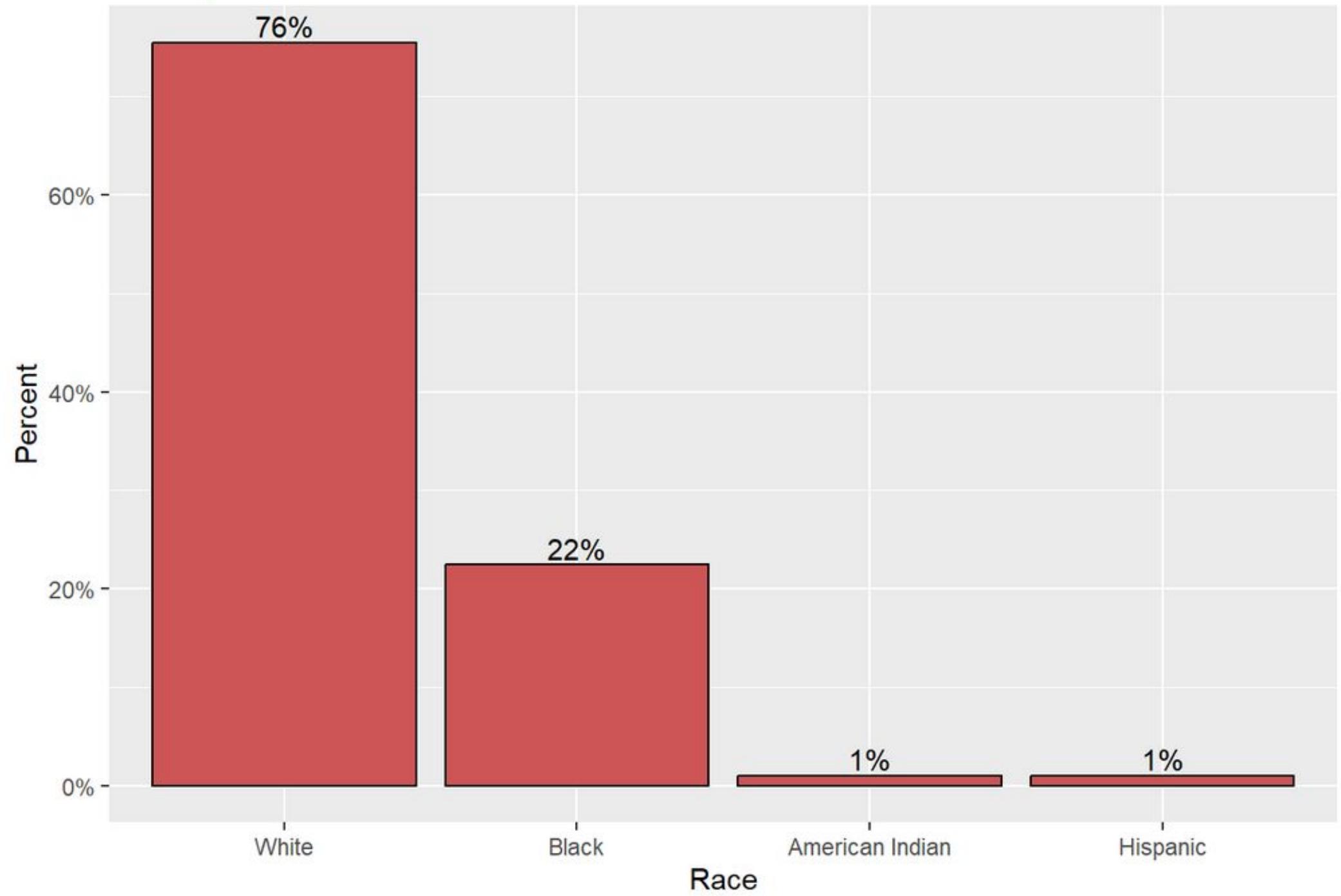
Bar Graph



Powered by

Futurense

Participants by race



- Can also be used to show percentage.
- Sorting in some order makes it more visually appealing.



Powered by

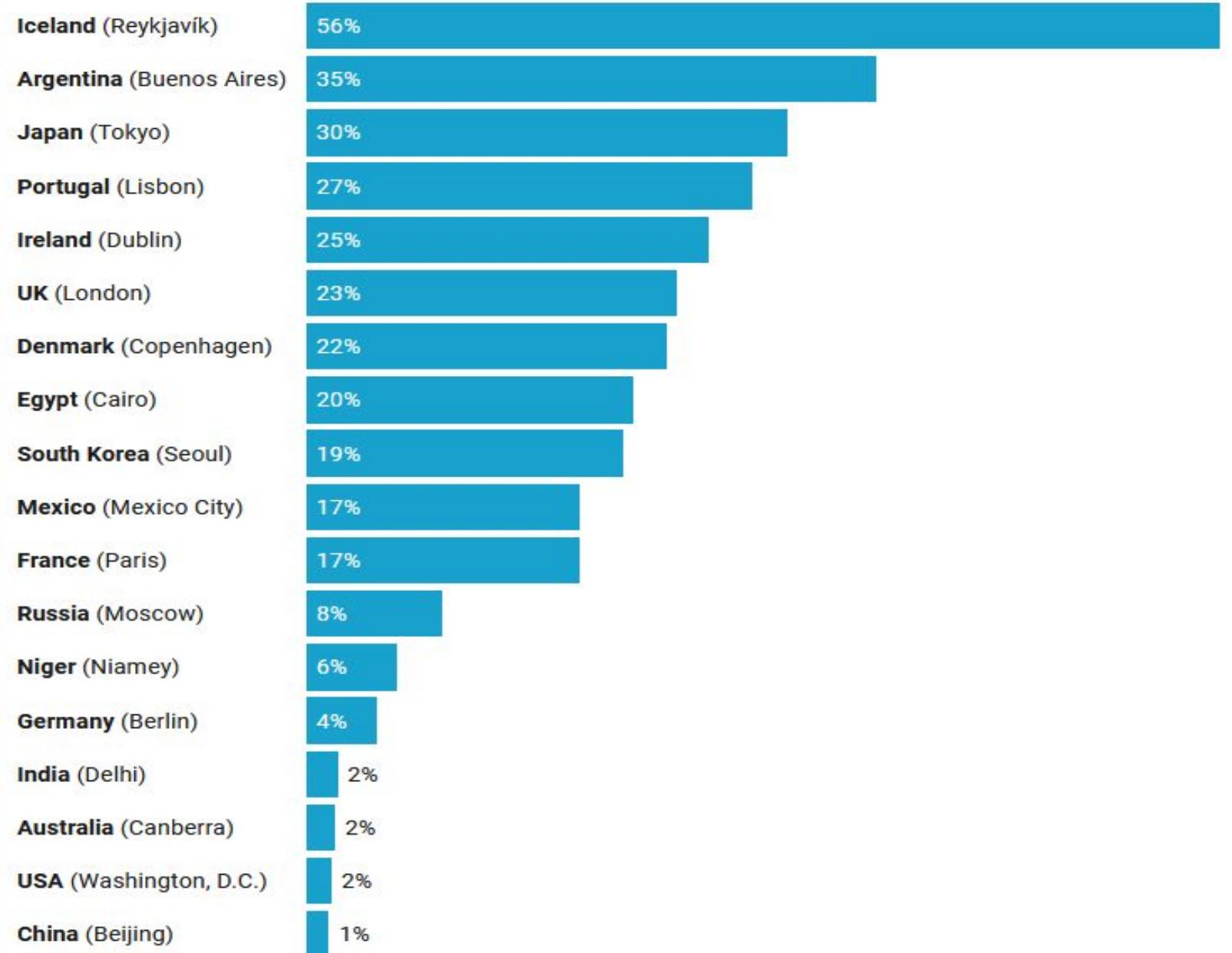


Futurense

Bar Chart

Rural and urban population

Share of population that lives in the capital, in urban areas and in rural areas, of selected countries, 2014.



The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

Source: [UN Population Division](#) • [Get the data](#) • Created with [Datawrapper](#)

BAR CHART WITH MANY BARS

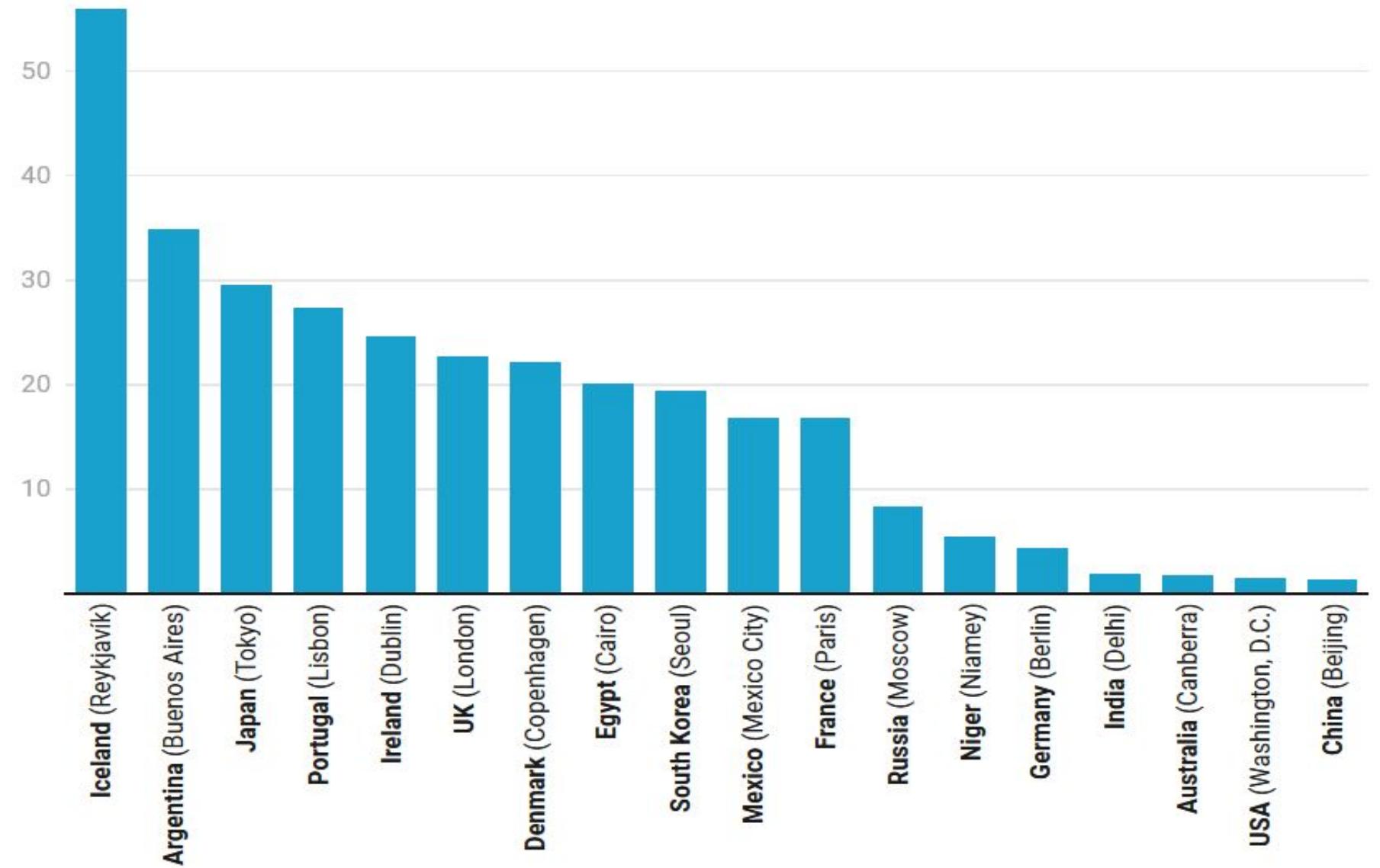
- Arranged horizontally



Column Chart

Rural and urban population

Share of population that lives in the capital, in urban areas and in rural areas, of selected countries, 2014.



The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

Source: [UN Population Division](#) • Get the data • Created with [Datawrapper](#)

COLUMN CHART

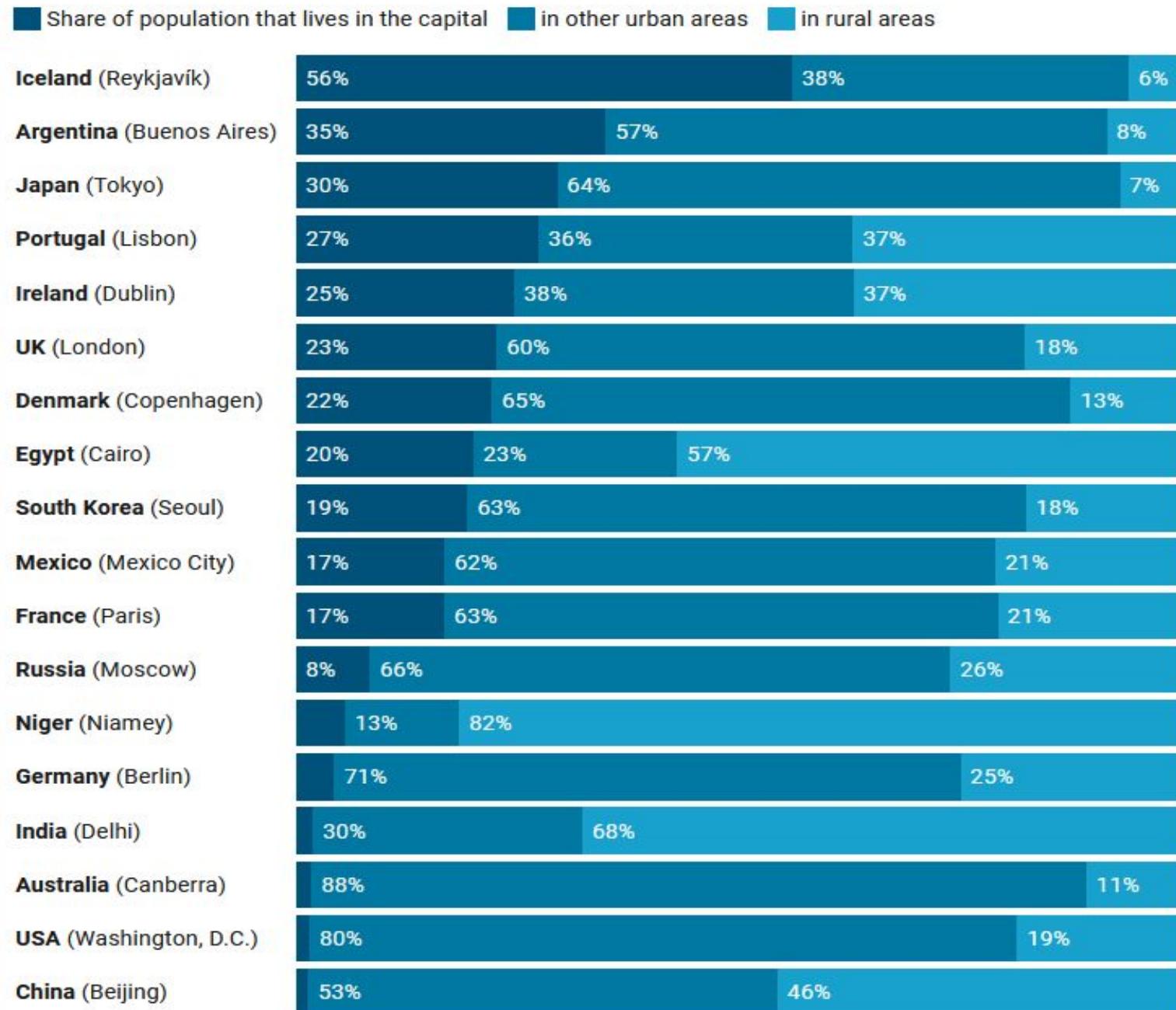
- Can be vertical
- Which one to choose can depend on the number of categories in the data.
- Can you fit them on one page or one slide?



Stacked Bar Chart

Rural and urban population

Share of population that lives in the capital, in urban areas and in rural areas, of selected countries, 2014.



The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

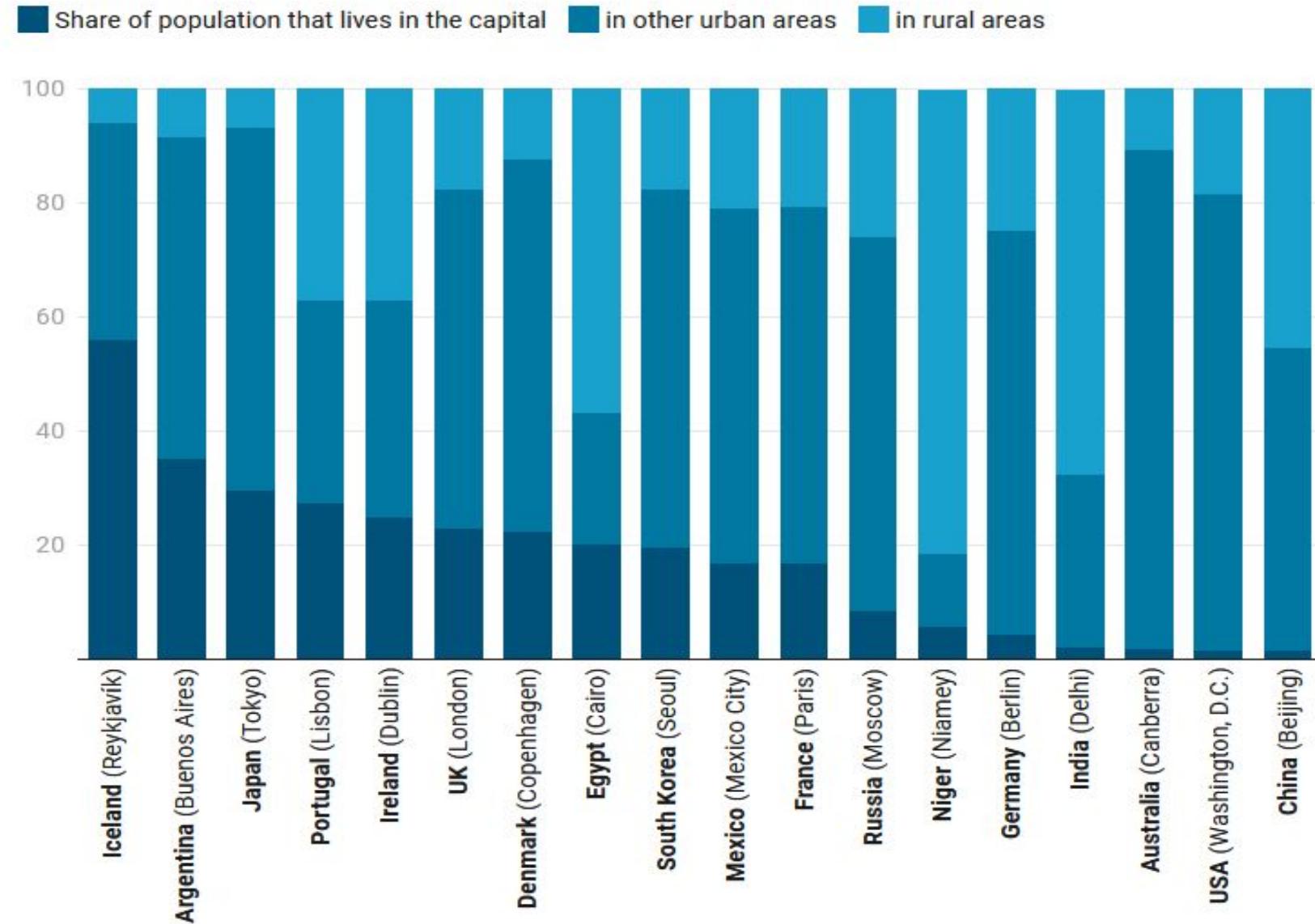
Source: UN Population Division • Get the data • Created with Datawrapper

- Illustrate different components of a variable for different categories.
- It shows the relationship between individual values and the total sum of the data points.
- Each bar consists of multiple segments stacked horizontally on top of each other.
- Can be hard to compare values across sub-categories since only the bottom segment starts from a consistent baseline.

Stacked Column Chart

Rural and urban population

Share of population that lives in the capital, in urban areas and in rural areas, of selected countries, 2014.



The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

Source: UN Population Division • Get the data • Created with Datawrapper

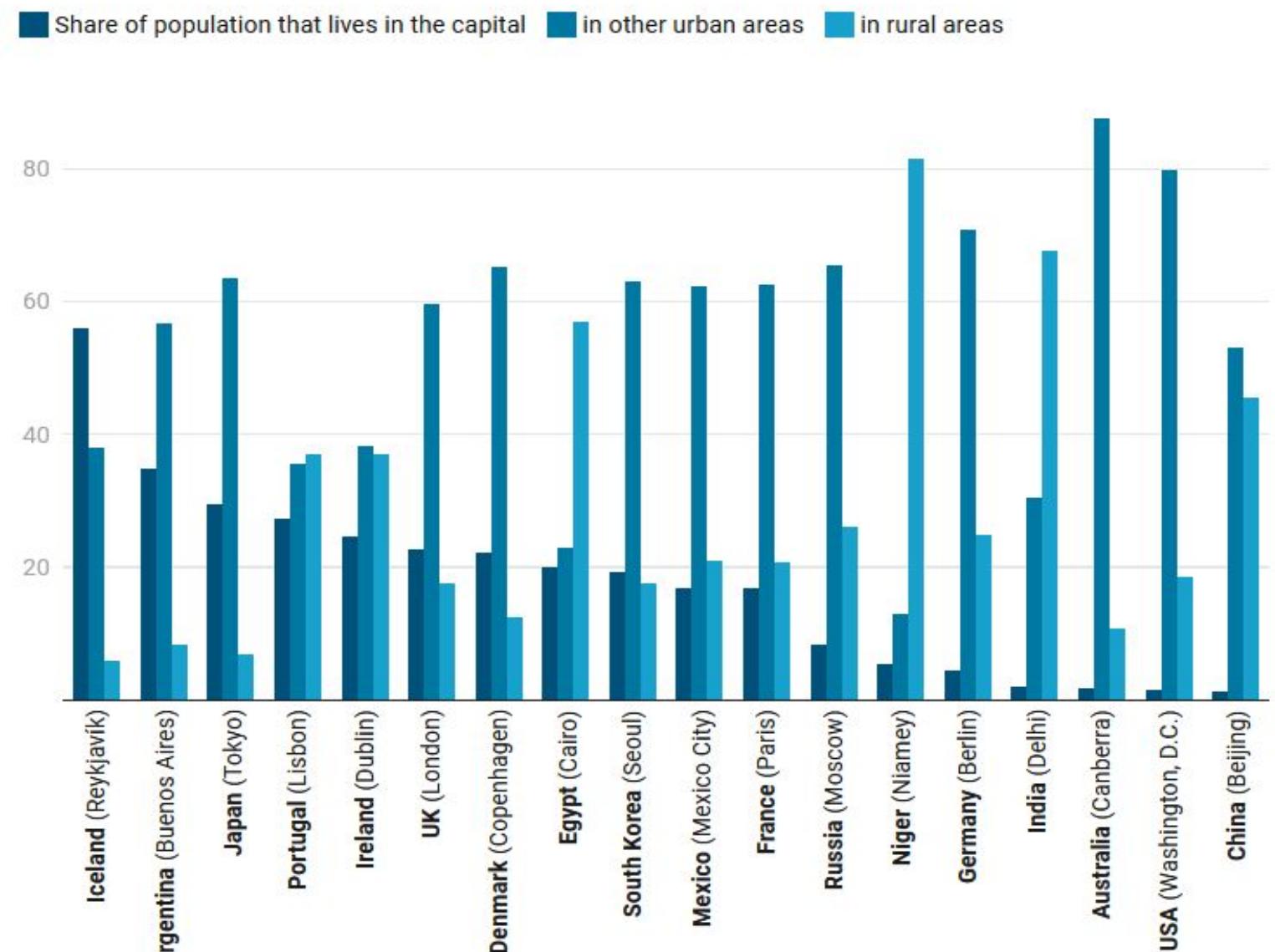
Can also be stacked vertically.



Grouped Column Chart

Rural and urban population

Share of population that lives in the capital, in urban areas and in rural areas, of selected countries, 2014.



The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

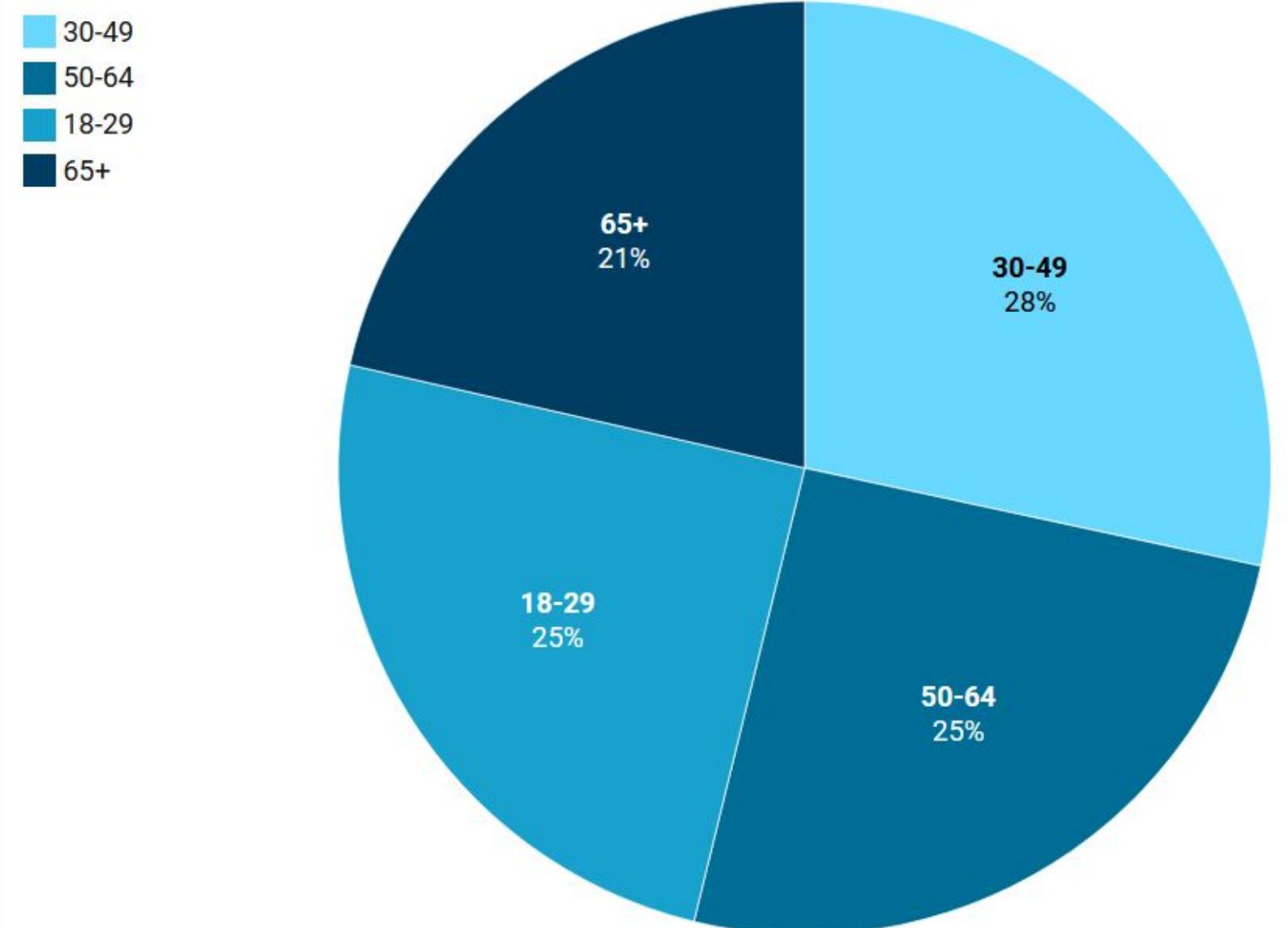
Source: [UN Population Division](#) • Get the data • Created with [Datawrapper](#)

- Organizes data into groups
- Easy comparison across sub-categories within each main category
- Displays each sub-category as an individual bar next to others in the same group
- Simple to visually compare values across multiple sub-categories within the same main category
- Can become cluttered and difficult to read when there are too many sub-categories or groups.

Pie Chart

Social media demographics

Percentage of U.S. adults who say they "ever use" selected social networks, by age group.



- Used to show the size of components of a whole
- Usually percentages
- Components need to add up to 100%

Survey of U.S. adults conducted Feb. 1-June 10, 2024.

Source: Pew Research • Get the data • Created with Datawrapper



Time Series Chart



Powered by
Futurense

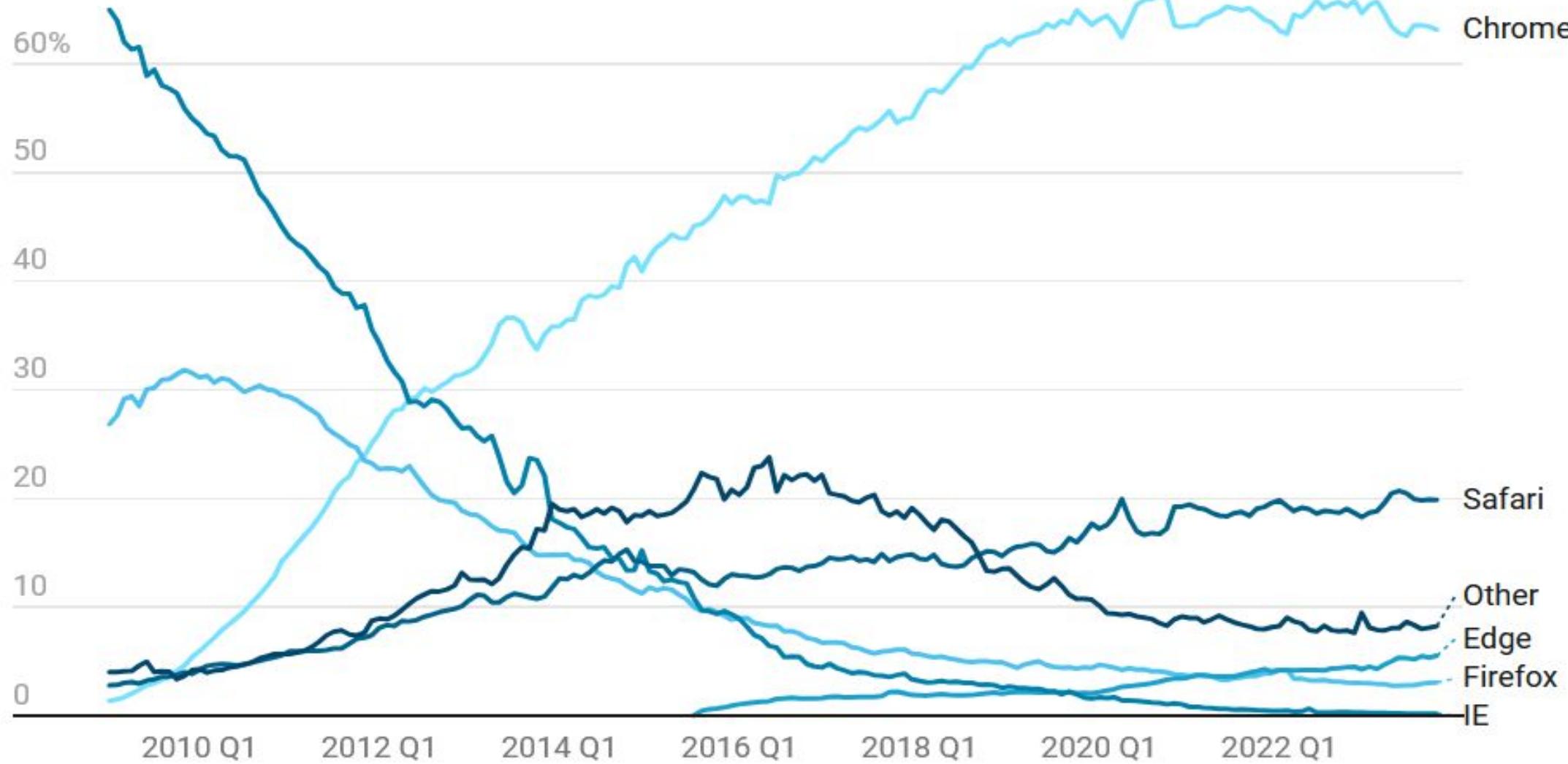


- A time series is a set of quantitative values obtained at successive time points
- A time series graph displays change over time.

Line Chart

The rise of Google Chrome

Web browser market share, January 2009 to October 2023



- Display multiple time series variables in one graph
- Can compare their performances over time
- Make sure they are over the same time frame

The UN defines "Urban" differently for each country. To count as "urban", Japanese settlements need to have at least 50,000 inhabitants. In Iceland, 200 inhabitants are enough.

Source: StatCounter GlobalStats • Get the data • Created with Datawrapper



Bivariate Visualization



Powered by

Futurense

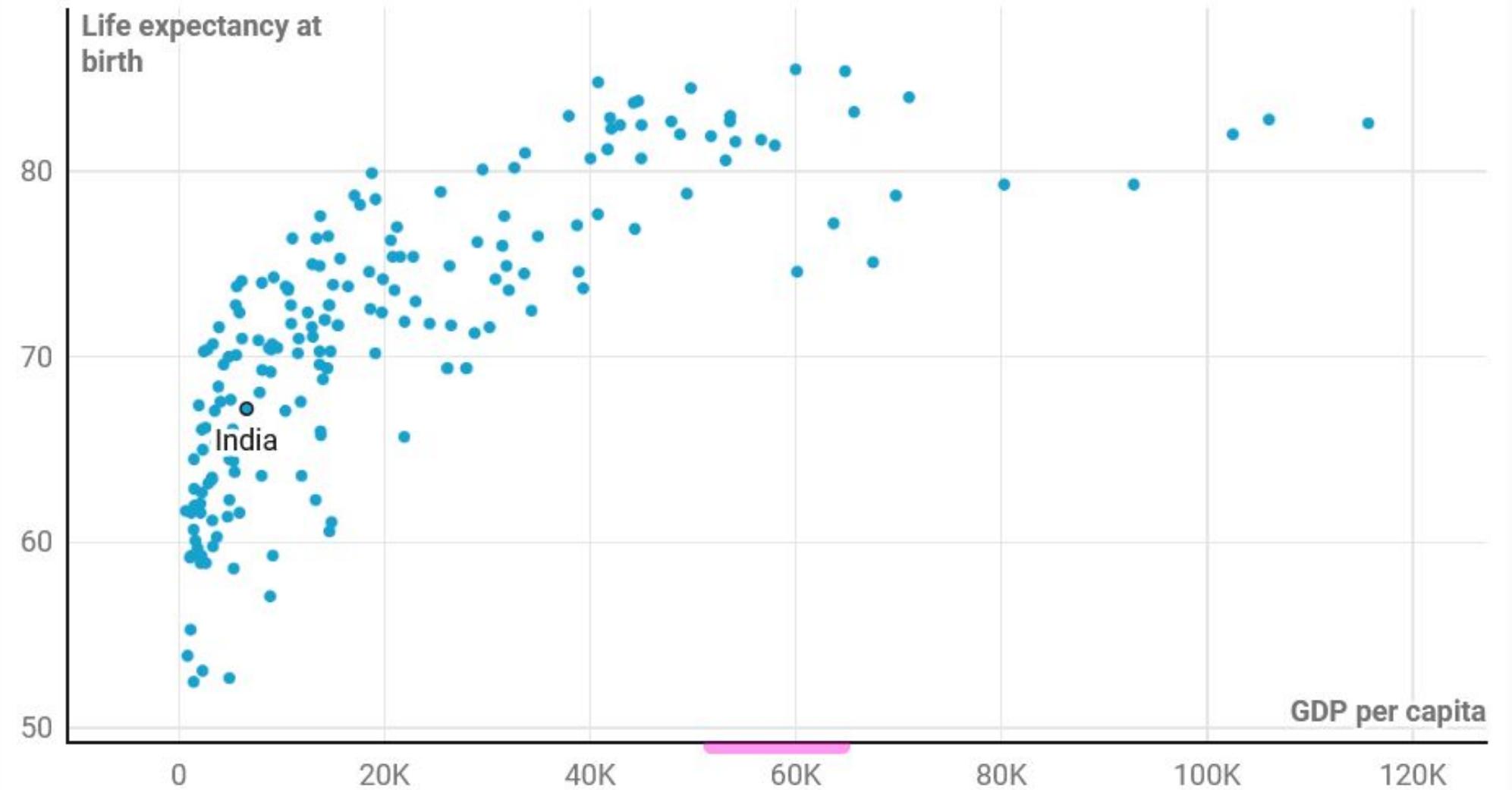
- After investigating each variable individually, we compare the relationship between two variables.
- Illustrate how two variables vary together

Scatter Plot

Income vs life expectancy

GDP per capita in US-Dollars and life expectancy in years for selected countries, 2021.

Country



GDP per capita, PPP in constant 2017 international \$

Source: Our World In Data • Get the data • Created with [Datawrapper](#)

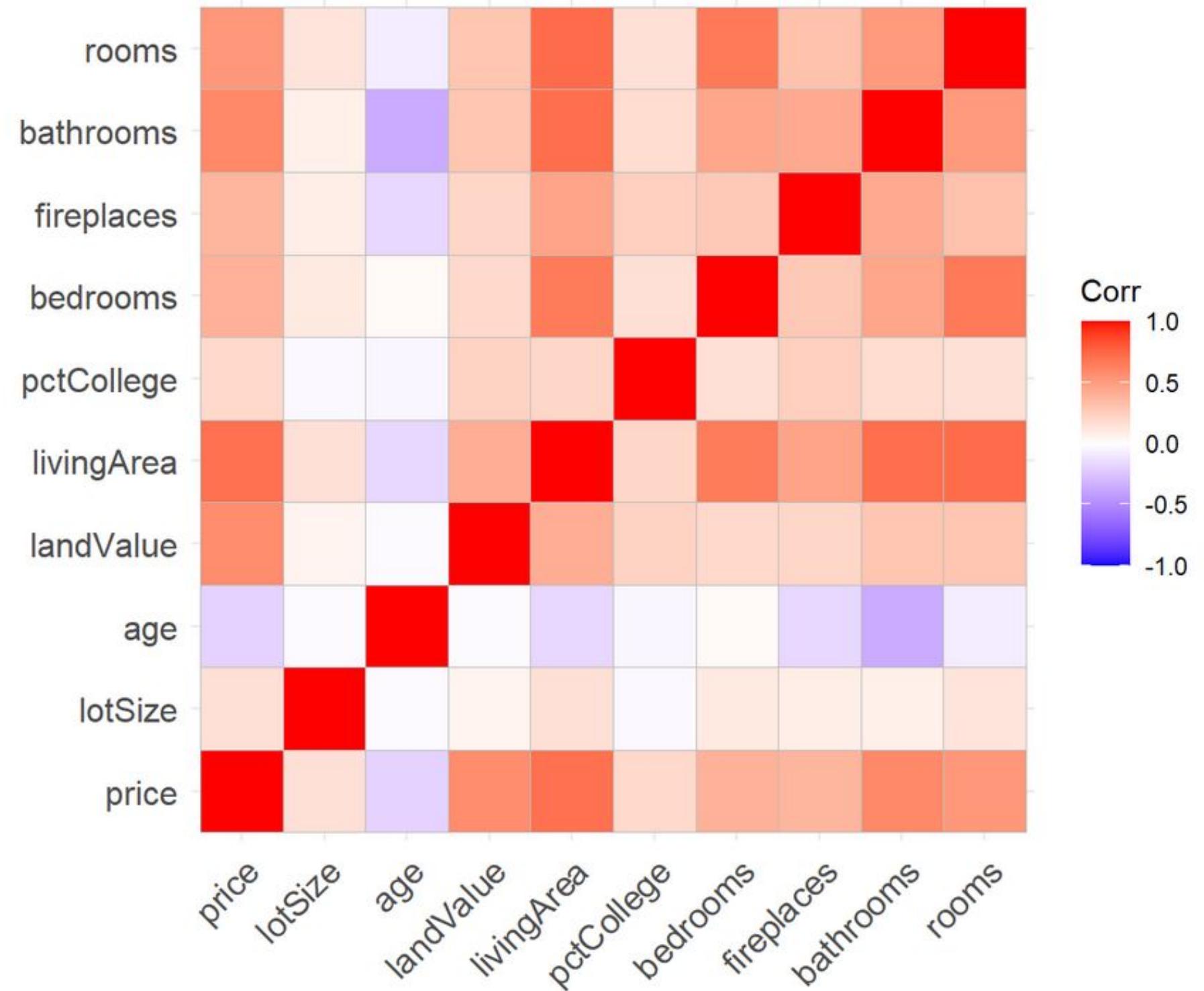
- Each point tracks the two values for a given record.
- Can mark an individual observation to draw the audience's attention.
- Use other aspects such as colours, texture to include more components.

Correlation Plot



Powered by

Futurense



- Visualize the pairwise relationships between a set of quantitative variables by displaying their correlations using color or shading.
- Can look at multiple pairs of variables at a time

Recap



Powered by
Futurense

Common visualizations



Powered by



Futurense

Thank you

